The EMBO Journal vol.8 no.9 pp.2601 – 2604, 1989

A collagen-binding 59-kd protein (fibromodulin) is structurally related to the small interstitial proteoglycans PG-S1 and PG-S2 (decorin)

Åke Oldberg, Per Antonsson, Karin Lindblom and Dick Heinegård

Department of Medical and Physiological Chemistry, Lund University, PO Box 94, S-221 00 Lund, Sweden

Communicated by R.Timpl

We have determined the primary structure of a 59 kd collagen binding protein which is present in many types of connective tissues, e.g. cartilage, tendon, skin, sclera and cornea. The amino acid sequence, deduced from a 2662 bp cDNA clone, predicts a 42 kd protein with a high content of leucine residues. Most of the protein consists of homologous 23 amino acid residues repeats with predominantly leucine residues in conserved positions. Similar leucine rich repeats have been identified in a number of proteins including the small interstitial proteoglycans decorin and PG-S1. The 59 kd protein and the two proteoglycans are homologous in their entire sequences suggesting that they have evolved from a common ancestral gene. The 59 kd protein and decorin are also functionally related in that both bind to collagen type I and II, and affect their fibrillogenesis. The substitution with glycosaminoglycan chains appears to be a feature shared by all three members of this family of leucine rich motif extracellular proteins, since the 59 kd protein isolated from cartilage is substituted with at least one keratan sulfate chain.

Key words: fibromodulin/proteoglycan/PG-S1/PG-S2/decorin

Introduction

Connective tissue matrices contain collagens, glycoproteins and proteoglycans as major components. The collagens, which vary between tissues, can be divided into subclasses with rather different properties. The major class represents the typical interstitial fiber forming collagens, i.e type I and type III in most tissues, and type II in cartilage. Most connective tissues contain two types of small proteoglycans in different proportions. Both decorin (Krusius and Ruoslahti, 1986) and PG-S1 (Fisher *et al.*, 1989) from human sources have been cloned and sequenced. They show extensive homologies, but clearly represent different gene products. Both proteins show internal 25 amino acid repeats. Most marked structural differences are observed in the structure of their N-terminal parts, where decorin has one chondroitin sulfate/dermatan sulfate side chain and PG-S1 has two glycosaminoglycan side chains.

Although tissues contain the same major constituents they are different in their structural appearance and mechanical properties. The supramolecular organization of collagen I, for instance, varies considerably between tissues, depending on tensile requirements. The mechanisms and factors involved in the process of collagen fibril formation are virtually unknown. Candidates as modulators of fibril assembly and structure are proteins that bind to collagen. Such proteins may also alter the surface properties e.g. charge of the collagen fiber. An example of such a macromolecule is the proteoglycan decorin, that binds to collagen type I and II and affects the rate of fibril formation *in vitro* (Vogel *et al.*, 1984). Fibrils formed in the presence of decorin are thinner than those formed without the proteoglycan (Vogel *et al.*, 1984). Similar inhibitory effects on collagen fibril formation *in vitro* are obtained in the presence of a 59 kd connective tissue protein (Hedbom and Heinegård, 1989). This protein was originally isolated form cartilage where it represents 0.1–0.3% of tissue wet weight (Heinegård *et al.*, 1986). It is also present in other tissues, e.g. sclera, tendon and cornea, and has a characteristic amino acid composition, where 14% of the residues are leucine. We propose the name fibromodulin (FM) for the 59 kd protein since it binds to collagen and delays fibril formation *in vitro*.

Results

Cloning of cDNA

Previous studies with radioimmunoassay has shown that FM is present in many connective tissues, although it is somewhat more abundant in cartilage. Therefore we screened a λgt11 cDNA library constructed from bovine tracheal chondrocyte mRNA with an antiserum against FM. Screening of 3×10^5 recombinants resulted in the isolation of clone λFM1. This clone had a 2.2 kbp cDNA insert, which hybridized to a 3 kb mRNA in transfer blot analysis of chondrocyte RNA. Partial sequence analysis indicated that λFM1 represented a 5′-truncated cDNA. To obtain longer cDNA clones, a restriction enzyme fragment corresponding to the 5′-end of λFM1 was radiolabeled and used to rescreen the library. Among the clones hybridizing to the probe we selected λFM2, having a 2.7 kbp insert for sequence analysis.

DNA and deduced amino acid sequence

The nucleotide sequence of λFM2 was determined for both strands using a combination of random and restriction enzyme fragmentation and specific priming with synthetic oligonucleotides (Figure 1). The 2662 bp cDNA contains a 1186 bp open reading frame followed by a 1486 bp non-coding region. The polyA sequence is preceded 16 bp upstream by an AATAAA polyadenylation signal. The first ATG in the DNA sequence presumably represents the translation initiation codon, since the deduced methionine residue is followed by a typical signal peptide sequence, consisting of hydrophobic and neutral amino acid residues. The signal peptide cleavage site, however, is unknown. Several attempts

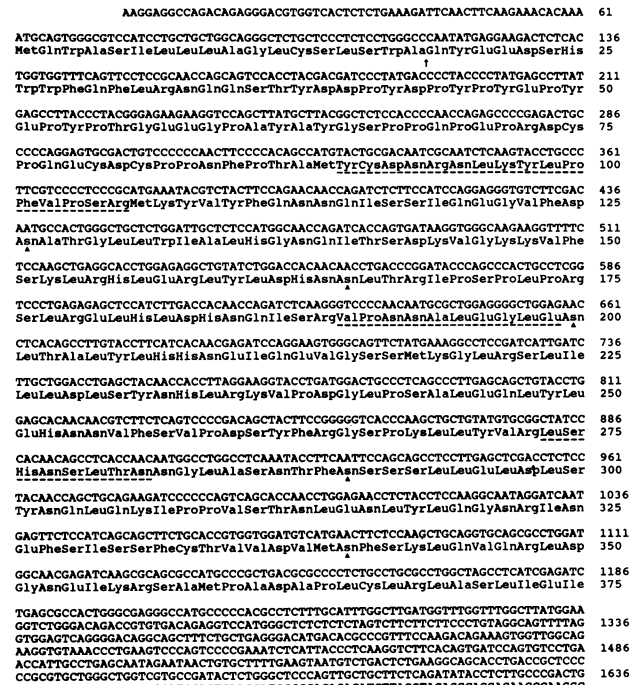

Fig. 1. The cDNA sequence and deduced amino acid sequence of bovine FM. Amino acid sequences confirmed by protein sequence analysis is underlined with a broken line. A putative signal peptide cleavage site is indicated with an arrow and possible *N*-glycosylation sites with (▲). The polyadenylation signal is underlined.

to determine the amino-terminal sequence of the purified protein have been unsuccessful, indicating a blocked N-terminus.

The identity of the deduced protein sequence was confirmed by amino acid sequence analysis of cyanogen bromide fragment and two tryptic peptides of FM. In further support the amino acid composition derived from the deduced sequence also agrees with the composition determined for FM. Most of the FM sequence, which contains 375 amino acid residues representing a molecular mass of 42 200 kd, consists of 10 repeats of ~23 residues. This region is located in the middle of the protein and is flanked on both sides by domains lacking the repeat structure. The amino terminal region presumably has a tertiary structure consisting of two loops resulting from disulfide bridges between the four cysteine residues. Similarly the carboxy-terminus may be a single loop as a result of a disulfide bridge between the two cysteine residues. The protein has five Asn-X-Ser/Thr sequences, which represent potential sites for *N*-glycosylation.

Homologies

The central domain in FM contains ten 23 residue long repeats with predominantly leucine residues in conserved positions (Figure 2). Similar leucine rich repeats are present in the small interstitial proteoglycans decorin and PG-S1 (Patthy, 1987; Ruoslahti, 1988; Fisher *et al.*, 1989). The repeat has also been found in three mammalian proteins; the serum protein LRG (Takashashi *et al.*, 1985), platelet

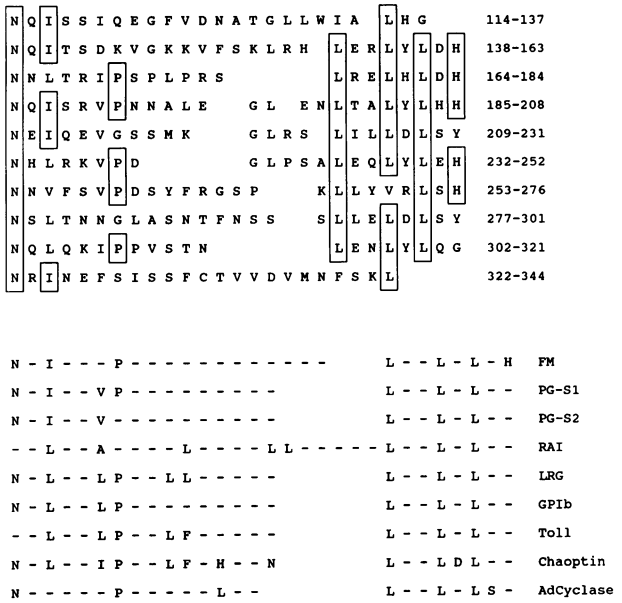

Fig. 2. Alignment of 10 leucine rich repeats representing amino acid residues 114–344 in FM. Identical residues are boxed. The consensus sequence of the leucine rich repeats in FM is compared to the consensus of similar repeats in other proteins.

surface protein GP1b (Lopez *et al.*, 1987) and a ribonuclease/angiogenin inhibitor (RAI) (Schneider *et al.*, 1988). Two *Drosophila* morphogenic proteins; chaoptin (Reinke *et al.*, 1988) and toll protein (Hashimoto *et al.*, 1988) contain the repeat, as does yeast adenylate cyclase (Kataoka *et al.*, 1985). Comparison of the entire FM sequence with other proteins show that FM is homolgous to the small proteoglycans (Figure 3). Decorin, PG-S1 and FM are similar in size and the cysteine residues are located in conserved positions.

Fibromodulin is substituted with keratan sulfate

We observed that ^{35}S-labeled FM could be immunoprecipitated from conditioned culture media of primary bovine sclera fibroblasts grown in the presence of [^{35}S]sulfate. To determine if the sulfate groups in the protein were substituents of glycosaminoglycan chains, we used the endoglycosidases chondroitinase ABC and keratanase. The chondroitin sulfate degrading enzyme did not reduce the size of cartilage FM as shown by SDS–PAGE (data not shown). Keratanase treatment of FM, on the other hand, resulted in a size reduction as demonstrated by SDS–PAGE (Figure 4). This shows that FM isolated from cartilage is substituted with keratan sulfate chain(s). Furthermore, a monoclonal antibody recognizing highly sulfated regions in keratan sulfate (Caterson *et al.*, 1983) bound to FM. The epitope recognized by this monoclonal antibody was destroyed upon keratanase digestion (Figure 4).

Discussion

The amino acid sequence of fibromodulin has been deduced from cloned cDNA. The sequence predicts a 42 kd protein containing 375 amino acid residues. The discrepancy between this molecular mass and the previously determined 59 kd by SDS–PAGE is probably due to anomalies in the electrophoretic behavior of FM as a result of carbohydrate substitution.

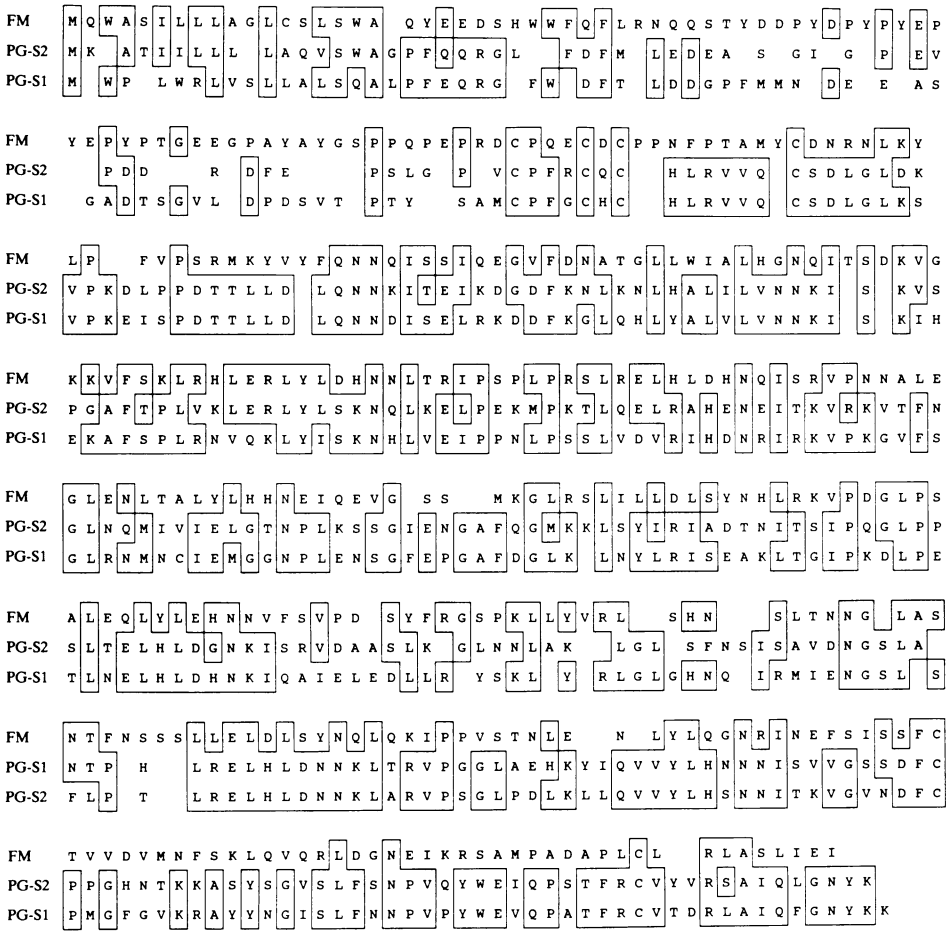

Fig. 3. Comparison of FM, PG-S1 and decorin. The sequences, including signal peptides, are aligned and identical residues are boxed. The PG-S1 and decorin sequences represent the proteins from human sources (Krusius and Ruoslahti, 1986; Fisher *et al.*, 1989).

The major part (60%) of FM consists of an ~23 amino acid residue long sequence repeated 10 times. This structure with predominantly leucine residues in conserved positions is also present in other proteins, ranging in origin from yeast to mammals. In most cases the repeats have been implicated in interactions. In yeast adenylate cyclase the repeats are located on the cytoplasmic side and has been postulated to anchor the enzyme to the cell membrane (Kataoka *et al.*, 1985). The *Drosophila* proteins chaoptin (Reinke *et al.*, 1988) and toll protein (Hashimoto *et al.*, 1988) play roles in the morphogenesis of photoreceptor cells and in the dorsal ventral embryonic patterning, respectively. It has been suggested in both cases that the repeat structures interact with the plasma membrane. The mechanisms for this interaction are not known, but it has been postulated that the leucine rich repeats in LRG, a serum protein of unknown function, form a secondary structure with hydrophobic residues oriented outwards and accessible for interaction with hydrophobic surfaces (Kataoka *et al.*, 1985). In the platelet surface protein GP1b the leucine repeat binds to von Willebrand factor and thrombin (Titani *et al.*, 1987). Also the ribonuclease/angiogenin inhibitor has been implicated in strong protein–protein interactions (Schneider *et al.*, 1988). In the proteoglycan decorin the repeat structure could be responsible for the binding to collagen (Vogel *et al.*, 1985)

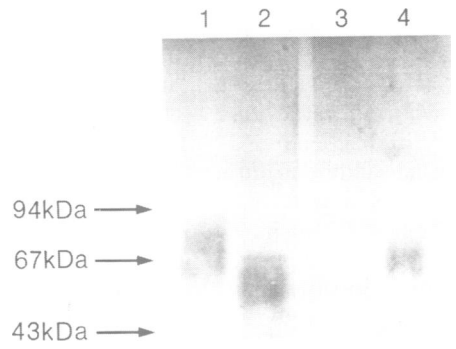

Fig. 4. Protein transfer blot analysis of FM digested with keratanase. Bovine cartilage FM was electrophoresed on a polyacrylamide gel in SDS before or after endoglycosidase digestion, transferred to introcellulose filters and identified with antibodies. **Lane 1:** FM incubated with FM antibodies. **Lane 2**: Keratanase digested FM incubated with FM antibodies. **Lane 3:** Keratanase digested FM incubated with keratan sulfate antibodies. **Lane 4:** FM incubated with keratan sulfate antibodies.

or to cell surfaces (Glössl *et al.*, 1983). The significance of the leucine repeat in FM is unknown, but the structural and functional relationship with decorin suggests a role in collagen binding.

FM, the proteoglycans decorin and PG-S1 are homologous, not only in the leucine rich repeat domain, but in their entire sequences. FM, decorin and PG-S1 are similar in size (357–375 amino acid residues) and have the same general structure with cysteine residues in conserved positions. This suggests that the three proteins have evolved from a common ancestral gene. Decorin and PG-S1, which contain 55% identical amino acid residues, differ in their amino terminal domains. These regions contain serine residues substituted with chrondroitin or dermatan sulfate. PG-S1 has two glycosaminoglycan chains, whereas decorin is substituted with a single chain. Also FM is substituted with glycosaminoglycan in the form of keratan sulfate. The keratan sulfate chain(s) will give FM polyanionic similar to decorin and PG-S1.

The structural similarities suggest that FM and the proteoglycans are also functionally related. Decorin binds to collagen type I and II (Vogel *et al.*, 1984) and fibronectin (Schmidt *et al.*, 1987). In *in vitro* assays decorin causes a delay in collagen fibril formation and the fibrils formed are thinner than those formed in the absence of decorin (Vogel *et al.*, 1984). Furthermore, proteoglycans appear to bind to collagen also *in vivo*, since staining for their polyanionic side chains in e.g. cornea shows location at the 'd' and 'e' bands in the collagen fiber (Scott and Haigh, 1988). Also FM binds to collagen type I and II and causes a slower rate of fibril formation *in vitro* (Hedbom and Heinegård, 1989). The FM binding site on the collagen molecule has not been determined but appears to be different from the decorin binding site since FM and decorin appear to have additive effects in the collagen fibrillation assay (Hedbom and Heinegård, 1989). Since FM and decorin are synthesized by the same cells in many connective tissues, they presumably bind to different sites on the collagen fibril and fulfill different functions in the formation of the collagen network. In this context it is interesting to note that keratan sulfate proteoglycans have tentatively been identified at the collagen fibril 'a' and 'c' band in cornea (Scott and Haigh, 1988). Possibly one of these proteoglycans is FM, since the protein is expressed in cornea, can be substituted with keratan sulfate and binds to collagen. In contrast to FM and decorin, PG-S1 has not been reported to bind to collagen, nor has this proteoglycan been shown to interact with any other extracellular matrix component. Extraction of PG-S1 from tissues, however, requires strong salt or denaturing solvents indicating binding to other matrix constituents.

In addition to its binding to collagen and fibronectin, decorin has been implicated in the regulation of cell proliferation, possibly by altering the deposition of extracellular matrix components (Yamaguchi and Ruoslahti, 1988). The synthesis of small proteoglycans is regulated by TGF-β and in some cell lines as much as a 20-fold increase in synthesis has been reported (Bassols and Massague, 1988). These observations support the notion that this family of leucine rich motif proteins may have important regulatory functions in the extra cellular matrix.

Materials and methods

Construction and screening of the cDNA library

RNA was isolated from bovine tracheal chondrocytes and used to construct a cDNA library in the expression vector λgt11 as previously described (Oldberg *et al.*, 1987). The library was screened with an antiserum prepared against bovine fibromodulin (Heinegård *et al.*, 1986). The methods used for RNA transfer blot analysis and DNA sequence analysis have been described (Oldberg *et al.*, 1986).

Fragmentation and sequence analysis of fibromodulin

Bovine fibromodulin was cleaved by treatment with 1% cyanogen bromide in formic acid. Trypsinized fragments were generated by treating fibromodulin with TPCK-treated trypsin (substrate to enzyme ratio of 100) in 1% ammonium bicarbonate for 5 h at 37°C (Oldberg *et al.*, 1988). Fragments were separated on serially coupled columns of Superose 6 and Superose 12 eluted with 4 M Guanidine–HCl and subjected to protein sequence analysis as previously described (Oldberg *et al.*, 1986).

Glycosaminoglycan analysis

Digestions of FM with keratanase (keratan sulfate 1,4-β-D-galactano hydrolase, EC 3.2.1.103) (Boehringer) was performed according to the manufacturers recommendation. A monoclonal antibody (5D4) specific to keratan sulfate was a gift from Dr B.Caterson, University of West Virginia, USA (Caterson *et al.*, 1983). SDS–PAGE was done according to Laemmli (1970). Proteins were transferred to nitrocellulose (Bowen *et al.*, 1980), reacted with antibodies to fibromodulin and to keratan sulfate, respectively, and detected with alkaline phosphatase conjugated second antibodies (Dakopatts, Denmark).

Acknowledgements

This work was supported by grants form the Swedish medical research Council, Folksams stiftelse, Konung Gustaf V:s 80-årsfond, Kock's stiftelse, Österlunds stiftelse and the medical faculty, University Lund. P.A. was supported by a grant from Tesdorpfs Stiftelse. We are indebted to Dr L.Fisher, National Institutes of Health, USA, for providing us with the sequence of PG-S1 prior to publication.

References

Bassols,A. and Massague,J. (1988) *J. Biol. Chem.*, **263**, 3030–3045.
Bowen,B., Steinberg,J., Laemli,U.K. and Weintraub,H. (1980) *Nucleic Acids Res.*, **8**, 1–20.
Caterson,B., Christner,J. and Baker,J.R. (1983) *J. Biol. Chem.*, **258**, 8848–8854.
Fisher,L.W., Termine,J.D. and Young,F.M. (1989) *J. Biol. Chem.*, **264**, 4571–4576.
Glössl,J., Schubert-Prinz,R., Gregory,J.D., Damble,S.P., von Figura,K. and Kresse,H. (1983) *Biochem. J.*, **215**, 295–301.
Hashimoto,C., Hudson,K.L. and Anderson,K.V. (1988) *Cell*, **52**, 269–279.
Hedbom,E. and Heinegård,D. (1989) *J. Biol. Chem.*, **264**, 6898–6905.
Heinegård,D., Larsson,T. and Sommarin,Y., Franzén,A., Paulsson,M. and Hedbom,E (1986) *J. Biol Chem.*, **261**, 13866–13872.
Kataoka,T., Brock,D. and Wigler,M. (1985) *Cell*, **43**, 493–505.
Krusius,T. and Ruoslahti,E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7683–7687.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Lopez,J.A., Chung,D.W., Fujikawa,K., Hagen,F.S., Papayannopoulou,T. and Roth,G.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5615–5619.
Oldberg,Å, Franzén,A. and Heinegård,D. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8819–8823.
Oldberg,Å, Antonsson,P. and Heinegård,D. (1987) *Biochem. J.*, **243**, 255–259.
Oldberg,Å, Franzén,A. and Heinegård,D. (1988) *J. Biol. Chem.*, **263**, 19430–19432.
Patthy,L. (1987) *J. Mol. Biol.*, **198**, 567–577.
Reinke,R., Krantz,D.E., Yen,D. and Zipursky,S.L. (1988) *Cell*, **52**, 291–301.
Ruoslahti,E. (1988) *Annu. Rev. Cell. Biol.*, **4**, 229–255.
Schmidt,G., Robenek,H., Harrach,B., Glössl,J., Nolte,V., Hörmann,H., Richter,H. and Kresse,H. (1987) *J. Cell. Biol.*, **104**, 1683–1691.
Schneider,R., Schneider-Scherzer,E., Thurnher,M., Auer,B. and Schweiger,M. (1988) *EMBO J.*, **7**, 4151–4156.
Scott,J.E. and Haigh,M. (1988) *Biochem. J.*, **253**, 607–610.
Takahashi,N., Takahashi,Y. and Putnam,F.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1906–1910.
Titani,K., Takio,K., Handa,M. and Ruggeri,Z.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5610–5614.
Vogel,K.G., Paulsson,M. and Heinegård,D. (1984) *Biochem. J.*, **223**, 587–597.
Yamaguchi,Y. and Ruoslahti,E. (1988) *Nature*, **336**, 244–246.

Received on May 11, 1989